Single-index Regression models with right-censored responses

Olivier Lopez

Single Index Regression Models with right censored responses

Olivier Lopez*

Crest-Ensai and Irmar

March 7, 2008

Abstract

In this article, we propose some new generalizations of M-estimation procedures for single-index regression models in presence of randomly right-censored responses. We derive consistency and asymptotic normality of our estimates. The results are proved in order to be adapted to a wide range of techniques used in a censored regression framework (e.g. synthetic data or weighted least squares). As in the uncensored case, the estimator of the single-index parameter is seen to have the same asymptotic behavior as in a fully parametric scheme. We compare these new estimators with those based on the average derivative technique of Burke and Lu (2005) through a simulation study.

Key words: semiparametric regression, dimension reduction, censored regression, Kaplan-Meier estimator, single-index models.

1 Introduction

In regression analysis, one investigates on the function $m(x) = E[Y \mid X = x]$, which is traditionally estimated from independent copies $(Y_i, X_i)_{1\leq i\leq n} \in \mathbf{R}^{1+d}$. The parametric

*Crest-Ensai and Irmar, rue Blaise Pascal, 35000 Bruz, France. E-mail : lopez@ensai.fr

approach consists of assuming that the function m belongs to some parametric family, that is $m(x) = f_0(\theta_0, x)$, where f_0 is a known function and θ_0 an unknown finite dimensional parameter. On the other hand, the nonparametric approach requires fewer assumptions on the model, since it consists of estimating m without presuming the shape of the function. However, this approach suffers from the so-called "curse of dimensionality", that is the difficulty to estimate properly the function m when the dimension d is high (in practice, $d \geq 3$). To avoid this important drawback of nonparametric approaches, while allowing more flexibility than a purely parametric model, one may use the semi-parametric single-index model (SIM in the following) which states

$$m(x) = E[Y \mid X'\theta_0 = x'\theta_0] = f(x'\theta_0; \theta_0),$$

where f is an unknown function and θ_0 an unknown finite dimensional parameter. If θ_0 were known, the problem would consist of a nonparametric one, but with the covariates belonging nevertheless to a one-dimensional space.

In this framework, numerous semi-parametric approach have been proposed for root-n consistent estimation of θ_0. Typically, these approaches can be split into three mains categories : M-estimation (Ichimura, 1993, Sherman, 1994b, Delecroix et Hristache, 1999, Xia et Li, 1999, Xia, Tong, et Li, 1999, Delecroix, Hristache et Patilea, 2006), average derivative based estimation (Powell, Stock et Stoker, 1989, Härdle et Stoker, 1989, Hristache et al., 2001a, 2001b), and iterative methods (Weisberg et Welsh, 1994, Chiou et Müller, 1998, Bonneu et Gba, 1998, Xia et Härdle, 2002).

If the responses of this regression model are randomly right-censored, these approaches clearly need to be adapted, for the random variable Y is not directly observed. The right censoring model states that, instead of observing Y, one observes i.i.d. replications of

$$\begin{aligned} T &= Y \wedge C, \\ \delta &= 1_{Y \leq C}, \end{aligned} \tag{1.1}$$

where C is some "censoring variable", and $\mathbf{1}_A$ denotes the indicator function of the set A. In this setting, semi-parametric Cox regression model (see e.g. Andersen et Gill, 1982) can be seen as a particular case of the SIM model, but allows less flexibility. Moreover, it is still interesting to extend mean-regression models to the censored framework. For this reason, Buckley and James (1978) proposed an estimator of the linear model under random censoring, and Lai and Ying (1991) and Ritov (1990) proved its asymptotic normality. Koul, Susarla and Van Ryzin (1981) initiated what we may call the "synthetic data" approach, based on transformations of the data. See Leurgans (1987), Zhou (1992b) and Lai & al. (1995). Zhou (1992a) also proposed a weighted least-square approach, applying weights in the least square criterion in order to compensate the censoring. These techniques were then used in the nonlinear regression setting, that is when f_0 is known but nonlinear. Stute (1999) established a connection between the weighted least-square criterion and Kaplan-Meier integrals. Delecroix, Lopez and Patilea (2006) extended the synthetic data approach. Heuchenne and Van Keilegom (2005) modified the Buckley-James' technique for polynomial regression purpose. When it comes to the SIM model under random censoring, Burke and Lu (2005) recently proposed an estimate using an extension of the average derivatives technique of Härdle and Stoker (1989) and the synthetic data approach of Koul, Susarla, Van Ryzin (1981).

In this paper, we propose a semi-parametric M-estimator of the SIM model under random censoring. We present a technique that is adapted to both main classes of censored regression techniques (synthetic data and weighted least squares), deriving root-n consistency of our estimate of θ_0, and then using it to estimate $m(x)$. Another advantage of our technique is that we do not require that the covariates X have a density with respect to Lebesgue's measure (only the linear combinations $\theta' X$ need to be absolutely continuous), which is an important advantage comparatively with the estimation procedure of Burke and Lu (2005).

The paper is organized as follows. In section 2 we present the regression model and our methodology. In section 3, we derive consistency of our semi-parametric estimates in a general form, asymptotic normality is obtained in section 4. A simulation study is presented in 5 to test the validity of our estimate with finite samples. Section 6 is devoted to technical proofs.

2 Model assumptions and methodology

In the following, we assume that we have the following regression model,

$$Y = f\left(\theta_0' X; \theta_0\right) + \varepsilon,$$

where θ_0 is a vector of first component equal to 1, and $E\left[\varepsilon \mid X\right] = 0$. The function f is defined in the following way, $f\left(u; \theta\right) = E\left[Y \mid X'\theta = u\right]$. Considering the censoring model (1.1), we will define the following distribution function,

$$\begin{aligned} F\left(t\right) &= \mathbb{P}\left(Y \leq t\right), \\ G\left(t\right) &= \mathbb{P}\left(C \leq t\right), \\ H\left(t\right) &= \mathbb{P}\left(T \leq t\right), \\ F_{(X,Y)}\left(x, t\right) &= \mathbb{P}\left(Y \leq t, X \leq x\right). \end{aligned}$$

In the following, we will assume that

$$\inf\{t, F(t) = 1\} = \inf\{t, H(t) = 1\}, \tag{2.2}$$

$$\mathbb{P}(Y = C) = 0. \tag{2.3}$$

Otherwise, if (2.2) does not hold, since some part of the distribution of Y remains unobserved, consistent estimation requires making additional restrictive assumptions on the law of the residuals. Note that, in this case, our estimators will still be root-n convergent, but not necessary to θ_0. Concerning (2.3), we use this assumption to avoid dissymetry problems between C and Y.

As a property of conditional expectation, for any function $J(\cdot) \geq 0$, we have

$$\begin{aligned} \theta_0 &= \arg\min_{\theta \in \Theta} E\left[\left(Y - f\left(\theta' X; \theta\right)\right)^2 J(X)\right] = \arg\min_{\theta \in \Theta} M\left(\theta, f\right) \qquad (2.4) \\ &= \arg\min_{\theta \in \Theta} \int \left(y - f\left(\theta' x; \theta\right)\right)^2 J(x) dF_{(X,Y)}\left(x, y\right). \end{aligned}$$

In equation (2.4), of course we can not exactly know θ_0, since two objects are missing in the definition of M, that is the distribution function $F_{(X,Y)}$ and the regression function $f\left(\theta' x; \theta\right)$. A natural way to proceed consists of estimating these two functions, and then plugging in these estimators into (2.4).

2.1 Estimating the distribution function

We already mentioned there are two main approaches for studying regression models in presence of censoring, the Weighted Least Square approach (WLS in the following) and the Synthetic Data approach (SD in the following).

The WLS approach. In the uncensored case, the distribution function $F_{(X,Y)}$ can be estimated using the empirical distribution. This tool is unavailable under random censoring, since it relies on the (unobserved) $(Y_i)_{1\leq i\leq n}$.Under random censoring, Stute (1993) proposed to use an estimator based on the Kaplan-Meier estimator of F. Recall the definition of Kaplan-Meier estimator,

$$\hat{F}(t) = 1 - \prod_{i:T_i\leq t}\left(1 - \frac{\sum_{j=1}^{n}\mathbf{1}_{\delta_j=1,T_j\leq T_i}}{1-\hat{H}(T_i-)}\right)^{\delta_i},$$

where $\hat{H}$ denotes the empirical distribution function of T. $\hat{F}$ can be rewritten as

$$\hat{F}(y) = \sum W_{in}\mathbf{1}_{T_i\leq y},$$

where W_{in} is the jump at observation i. It is particularly interesting to notice that the jump at observation i is connected to the Kaplan-Meier estimate of G at the same value (see, for example, Satten and Datta, 2000), that is

$$W_{in} = \frac{1}{n}\frac{\delta_i}{1-\hat{G}(T_i-)}. \tag{2.5}$$

Kaplan-Meier estimate is known to be a consistent estimate of F under the two following identifiability assumptions, that is

Assumption 1 Y *and* C *are independent.*

Assumption 2 $\mathbb{P}(Y\leq C \mid X,Y) = \mathbb{P}(Y\leq C\mid Y)$.

A major case for which Assumptions 1-2 hold is the case where C is independent from (Y,X). However, Assumption 2 is more general and covers a significant amount of situations (see Stute, 1999).

The SD approach. The SD approach consists of considering some alternative variable which has the same conditional expectation as Y. For this, observe that, through elementary calculus, under Assumptions 1-2,

$$\forall \phi,\ E\left[\frac{\delta\phi(X,T)}{1-G(T-)} \mid X\right] = E\left[\phi(X,Y) \mid X\right]. \tag{2.6}$$

From (2.6), we see that, if we define, accordingly to Koul & al. (1981),

$$Y^* = \frac{\delta T}{1-G(T-)},$$

we have $E[Y^* \mid X] = E[Y \mid X]$ under Assumption 1 and 2. Hence, if Y^* were available, the same regressions techniques as in the uncensored case could be applied to Y^*. Of course, Y^* can not be computed, since it depends on the unknown function G. But Y^* can be easily estimated (which is not the case for Y) by replacing G by its Kaplan-Meier estimate. For $i = 1, ..., n$ we obtain

$$\hat{Y}_i^* = \frac{\delta_i T_i}{1-\hat{G}(T_i-)}.$$

See also Leurgans (1987), Lai & al. (1995) for other kind of transformations.

Back to equation (2.4), the SD approach will first consists of observing that

$$\begin{aligned} \theta_0 &= \arg\min_{\theta\in\Theta} E\left[\left(Y^* - f(\theta' x;\theta)\right)^2 J(X)\right] = M^*(\theta, f) \qquad (2.7) \\ &= \arg\min_{\theta\in\Theta} \int \left(y^* - f(\theta' x;\theta)\right)^2 J(x) dF^*_{(X,Y^*)}(x, y^*), \end{aligned}$$

where $F^*_{(X,Y^*)}(x,y^*) = \mathbb{P}(X \le x, Y^* \le y^*)$.

Note that M^* and M are not the same functions. Indeed, Y^* happens to have the same conditional expectation as Y (hence M and M^* have the same minimizer θ_0), but it has not the same law.

2.2 Estimating $f(\theta' x;\theta)$

In the uncensored case, a common non-parametric way to estimate a conditional expectation is to use kernel smoothing. In this case, the Nadaraya-Watson estimate for $f(\theta' x;\theta)$ is

$$\begin{aligned} \hat{f}(\theta' x;\theta) &= \frac{\sum_{i=1}^n K\left(\frac{\theta' X_i - \theta' x}{h}\right) Y_i}{\sum_{i=1}^n K\left(\frac{\theta' X_i - \theta' x}{h}\right)} \\ &= \frac{\int y K\left(\frac{\theta' u - \theta' x}{h}\right) d\hat{F}_{emp}(u,y)}{\int K\left(\frac{\theta' u - \theta' x}{h}\right) d\hat{F}_{emp}(u,y)}. \end{aligned}$$

We are still facing the same problem of absence of the empirical distribution function. However, WLS and SD approaches can be used to extend the Nadaraya-Watson estimate to censored regression. In the following, we will only use the SD approach of Koul & al. to estimate the conditional expectation, that is

$$\hat{f}\left(\theta' x;\theta\right)=\frac{\sum_{i=1}^{n}K\left(\frac{\theta' X_i-\theta' x}{h}\right)\hat{Y}_i^*}{\sum_{i=1}^{n}K\left(\frac{\theta' X_i-\theta' x}{h}\right)}. \tag{2.8}$$

While using this estimator, we do not have to deal with Kaplan-Meier integrals at the denominator. In fact, the integral at the denominator becomes an integral with respect to the empirical distribution function of X. However, alternative estimates (not necessarily kernel estimates) can still be used, provided that they satisfy some further discussed conditions to achieve asymptotic properties of $\hat{\theta}$. Therefore we chose to present our results without presuming on the choice of $\hat{f}\left(\theta' x;\theta\right)$, and then to check in the Appendix section that the estimator defined in (2.8) satisfies the proper conditions.

Also observe that, using this kernel estimate, contrary to the average derivative technique of Burke and Lu (2005), we do not need to impose that X has a density with respect to Lebesgue's measure. We only need that the linear combinations $\theta' X$ do.

The choice of the trimming function J**.** The reason why we introduced the function J in (2.4) appears in the definition (2.8). To ensure uniform consistency of this estimate, we will need to bound the denominator away from zero. For this, we will need to restrain the integration domain to a set where $f_{\theta' X}(u)$ is bounded away from zero, $f_{\theta' X}$ denoting the density of $\theta' X$. If we were to know θ_0, we could consider a set $B_0=\{u: f_{\theta_0' X}(u)\geq c\}$ for some constant $c>0$, and use the trimming $J(\theta_0' X)=\mathbf{1}_{\theta_0' X\in B_0}$. Of course, this ideal trimming can not be computed, since it depends on the unknown parameter θ_0. Delecroix & al. (2006) proposed a way to approximate this trimming from the data. Given some preliminary consistent estimator θ_n of θ_0, they use the following trimming,

$$J_n(\theta_n' X)=\mathbf{1}_{\hat{f}_{\theta_n' X}(\theta_n' X)\geq c}.$$

In the following proofs, we will mostly focus on the estimation using the uncomputable trimming $J(\theta_0' X)$, and we will show in the appendix section that there is no asymptotic difference in using $J_n(\theta_n' X)$ rather than $J(\theta_0' X)$.

2.3 Estimation of the single-index parameter

Preliminary estimate of θ_0. For a preliminary estimate, we assume, as in Delecroix & al. (2006) that we know some set B such as $\inf_{x\in B,\theta\in\Theta}\{f_{\theta' X}(\theta' x) \geq c > 0\}$, and we consider the trimming function $\tilde{J}(x) = \mathbf{1}_{x\in B}$. To compute our estimate θ_n, we then can use either of the WLS or SD approach. For example, using the WLS approach, let

$$\theta_n = \arg\min_{\theta\in\Theta} \int \left(y - \hat{f}\left(\theta' x; \theta\right)\right)^2 \tilde{J}(x) d\hat{F}_{(X,Y)}\left(x, y\right) = \arg\min_{\theta\in\Theta} M_n^p(\theta, \hat{f}). \tag{2.9}$$

Estimation of θ_0. In view of (2.4) and (2.7), we will define our estimates of θ_0 according to the two regression approaches discussed above,

$$\begin{aligned}
\hat{\theta}_{WLS} &= \arg\min_{\theta\in\Theta_n} \int \left[y - \hat{f}\left(\theta' x; \theta\right)\right]^2 J_n(\theta_n' x) d\hat{F}_{(X,Y)}\left(x, y\right) \\
&= \arg\min_{\theta\in\Theta_n} M_n^{WLS}\left(\theta, \hat{f}\right), \\
\hat{\theta}_{SD} &= \arg\min_{\theta\in\Theta_n} \int \left[y^* - \hat{f}\left(\theta' x; \theta\right)\right]^2 J_n(\theta_n' x) d\hat{F}^*\left(x, y^*\right) \\
&= \arg\min_{\theta\in\Theta_n} M_n^{SD}\left(\theta, \hat{f}\right).
\end{aligned}$$

In the definition above, for technical convenience, we restrained our optimization to shrinking neighborhoods Θ_n of θ_0, chosen accordingly to the preliminary estimation by θ_n.

2.4 Estimation of the regression function

With at hand a root-n consistent estimate of θ_0, it is possible to estimate the regression function by using $\hat{\theta}$ and some estimate $\hat{f}$. For example, using $\hat{f}$ defined in (2.8) will lead to

$$\hat{f}\left(\hat{\theta}' x; \theta\right) = \frac{\sum_{i=1}^n K\left(\frac{\theta' X_i - \hat{\theta}' x}{h}\right) \hat{Y}_i^*}{\sum_{i=1}^n K\left(\frac{\hat{\theta}' X_i - \hat{\theta}' x}{h}\right)}.$$

3 Consistent estimation of θ_0

In this section, we prove consistency of θ_n where θ_n is defined in (2.9). As a consequence, $\hat{\theta}$ is consistent since it is obtained from minimization other a shrinking neighborhood of θ_0. We will need two kinds of assumptions to achieve consistency : general assumptions on the regression model including identifiability assumptions for θ_0, and assumptions on $\hat{f}$.

Identifiability assumptions for θ_0 and assumptions on the regression model.

Assumption 3 *$EY^2 < \infty$.*

Assumption 4 *If $M(\theta_1, f) = M(\theta_0, f)$, then $\theta_1 = \theta_0$.*

Assumption 5 *Θ and $\mathcal{X} = Supp(X)$ are compact subsets of $\mathbb{R}^d$ and f is continuous with respect to x and θ. Furthermore, assume that $|f(\theta_1' x; \theta_1) - f(\theta_2' x; \theta_2)| \leq \|\theta_1 - \theta_2\|^\gamma \Phi(X)$, for a bounded function $\Phi(X)$, and for some $\gamma > 0$.*

Assumption 3 is implicitly needed in order to define M, while Assumption 4 ensures the identification of the parameter θ_0. On the other hand, Assumption 5 states that the class of functions $\mathcal{F} = \{f(\theta'.; \theta), \theta \in \Theta\}$ is sufficiently regular to allow it to satisfy an uniform law of large numbers property. More precisely, Assumption 5 ensures that this class is Euclidean for a bounded envelope, according to Pakes and Pollard (1989). Observe that the condition that Φ is bounded can be weakened, by replacing it by a moment assumption on Φ. However, this condition is quite natural in a context where we will assume that the covariates are bounded random vectors, and this will simplify our discussion. Moreover, it implies that f is a bounded function of θ and x.

Assumptions on $\hat{f}$.

Assumption 6 *For all function g, define, for $c > 0$,*

$$\|g\|_\infty = \sup_{\theta \in \Theta, x} |g(\theta' x; \theta)| \mathbf{1}_{f_\theta(\theta' x) > c/2}.$$

Assume that $\|\hat{f} - f\|_\infty = o_P(1)$.

See section 6 for more details to see that the kernel estimator (2.8) satisfies this assumption under some additional integrability assumptions on the variable Y.

Theorem 3.1 *Under Assumptions 3 to 6, we have*

$$\sup_{\theta \in \Theta} \left| M_n\left(\theta, \hat{f}\right) - M_\infty(\theta, f) \right| = o_P(1).$$

As an immediate corollary, in a probability sense, $\theta_n \to \theta_0$.

Proof.

Step 1 : replacing $\hat{f}$ by f. Observe that, since the integration domain is restricted to the set B,

$$\begin{aligned}\left|M_n(\theta, f) - M_n(\theta, \hat{f})\right| &\leq \|\hat{f} - f\|_\infty \\ &\times[\|\hat{f} + f\|_\infty \int d\hat{F}_{(X,Y)}(x,y) + 2\int |yd\hat{F}_{(X,Y)}(x,y)|].\end{aligned}$$

Now using Assumption 6, deduce that $\sup_{\theta\in\Theta}|M_n(\theta, f) - M_n(\theta, \hat{f})| = o_P(1)$.

Step 2 : $M_n(\theta, f)$. Showing that $\sup_{\theta\in\Theta}|M_n(\theta, f) - M(\theta, f)| = o_P(1)$ can then be done in the same way as in a nonlinear regression model such as in Stute (1999). See the proof of Theorem 1.1 in Stute (1999). ■

4 Asymptotic normality

As in the uncensored case, we will show that, asymptotically speaking, our estimators behave as if the true family of functions f were known. Hence studying the asymptotic normality of our estimates reduces to study asymptotic properties of estimators in a parametric censored nonlinear regression model, such as those studied by Stute (1999) and Delecroix & al. (2008). We first recall some elements about the case "f known" (which corresponds to a nonlinear regression setting), and then show that, under some additional conditions on $\hat{f}$ and on the model, our estimation of θ_0 is asymptotically equivalent to the one performed in this unreachable parametric model.

4.1 The case f known

This case can be studied using the results of Stute (1999) for the WLS approach, or the results of Delecroix & al. (2008) for the SD approach. We recall some assumptions under which the asymptotic normality of the corresponding estimators is obtained.

Assumptions on the model. We denote by $\nabla_\theta f(x;\theta)$ the vector of partial derivatives of $(x,\theta) \to f(\theta' x;\theta)$ with respect to θ, and $\nabla^2_\theta f$ the corresponding Hessian matrix.

Assumption 7 *$f(\theta' x; \theta)$ is twice continuously differentiable with respect to θ, and $\nabla_\theta f$ and $\nabla_\theta^2 f$ are bounded as functions of x and θ.*

Assumptions on the censoring. We need some additional integrability condition. We first need a moment assumption which is related to the fact that we need to have $E[Y^{*4}] < \infty$.

Assumption 8

$$\int \frac{y^4 dF(y)}{[1 - G(y-)]^3} < \infty.$$

Actually x is not involved in Assumption 8 as it is assumed to be bounded. Furthermore, in the case f known, this assumption can be weakened, but it will be needed in the case f unknown to obtain uniform consistency rate for $\hat{f}$. The following assumption is used in Stute (1995, 1996) to achieve asymptotic normality of Kaplan-Meier integrals.

Assumption 9 *Let*

$$C(y) = \int_{-\infty}^{y} \frac{dG(s)}{\{1 - H(s)\}\{1 - G(s)\}}.$$

Assume that

$$\int y C^{1/2}(y) dF_{(X,Y)}(x, y) < \infty.$$

See Stute (1995) for a full discussion on this kind of assumption. Using our kernel estimator for estimating the conditional expectation will lead us to a slightly stronger assumption (see the appendix section), which is

Assumption 10 *For some $\varepsilon > 0$,*

$$\int y C^{1/2+\varepsilon}(y)[1 - G(y-)]^{-1} dF_{(X,Y)}(x, y) < \infty.$$

In the following, we will use the (slightly) stronger Assumption 10 since it may simplify some proofs (see Lemma 6.2 and the proof of Theorem 4.1). However, Assumption 10 could be replaced by Assumption 9 if we were to use an estimator (not necessarly kernel estimator) which would not require Assumption 10 to satisfy the proper convergence assumptions. Note that this kind of assumption is classical in studying regression models with censored responses. Although it is not mentioned in Burke and Lu (2005), a similar assumption is implicitly needed to obtain equation (2.29) of Lai & al. (1995). In their proof of Lemma A.7 page 199 of Burke & al. (2005), the authors refer to equation (2.29) page 275 of Lai & al. (1995): this only holds under the condition C3 of Lai & al. (1995) which basically controls the tail behavior of the distributions.

The following Theorem can be deduced from the proof of Theorem 1.2 in Stute (1999) and of Theorem 4 in Delecroix & al. (2008). However, to make this article self-contained, a short proof of this result is postponed at section 6.1 of the appendix.

Theorem 4.1 *Define*

$$\psi(y,T,\delta)=\left[\frac{(1-\delta)\,\mathbf{1}_{T>y}}{1-H\,(T-)}-\int\frac{\mathbf{1}_{T>y,y>v}dG\,(v)}{\left[1-H\,(v)\right]^2}\right]$$

and let

$$\begin{aligned}
U^{WLS} &= \frac{\delta\left(T-f\left(\theta_0'X;\theta_0\right)\right)}{1-G\,(T-)}+\int\{y-f\left(\theta_0'x;\theta_0\right)\}V(y,T,\delta)dF_{(X,Y)}\left(x,y\right), \\
U^{SD} &= \left[\frac{\delta T}{1-G\,(T-)}-f\left(\theta_0'X;\theta_0\right)\right]+\int yV(y,T,\delta)dF_{(X,Y)}\left(x,y\right),
\end{aligned}$$

and let $W^{WLS}=E\left[\left(U^{WLS}\right)^2\right]$ *and* $W^{SD}=E\left[\left(U^{SD}\right)^2\right]$. *Let* M_n *and* M_∞ *denote respectively either* M_n^{WLS} *and* M, *or* M_n^{SD} *and* M^*. *We have, under Assumptions 1 to 10,*

$$M_n\left(\theta,f\right)=M_\infty\left(\theta,f\right)+O_P\left(\frac{\|\theta-\theta_0\|}{\sqrt{n}}\right)+o_P\left(\|\theta-\theta_0\|^2\right)+R_n, \tag{4.10}$$

$$M_n\left(\theta,f\right)=\frac{1}{2}\left(\theta-\theta_0\right)'V\left(\theta-\theta_0\right)+\left(\theta-\theta_0\right)'\frac{W_n}{\sqrt{n}}+o_P\left(n^{-1}\right)+R_n, \tag{4.11}$$

where R_n *does not depend on* θ, *where*

$$V = E\left[\nabla_\theta f\left(X;\theta_0\right)\nabla_\theta f\left(X;\theta_0\right)'\right],$$

and where $W_n\Longrightarrow\mathcal{N}(0,W)$, *for* $W=W^{WLS}$ *and* $W=W^{SD}$ *in the* $WLS-$*case and* $SD-$*case respectively.*

In view of Theorem 1 and 2 of Sherman (1994), (4.10) states that, in the case where f is known, $|\hat{\theta}-\theta_0|=O_P\left(n^{-1/2}\right)$, while (4.11) gives the asymptotic law of $\hat{\theta}$, showing that $n^{1/2}(\hat{\theta}-\theta_0)\Longrightarrow\mathcal{N}(0,V^{-1}WV^{-1})$, in both WLS and SD cases.

4.2 The case f unknown

As f is unknown in the SIM model, we need to add some conditions about the rate of convergence of $\hat{f}$.

Assumptions on f. If we evaluate the function $\nabla_\theta f(x;\theta)$ at the point (x,θ_0), a direct adaptation of Lemma A.5 of Dominitz and Sherman (2003) shows that

$$\nabla_\theta f(x;\theta_0) = f'(\theta_0' x)\{x - E[X \mid \theta_0' X = \theta_0' x]\}, \tag{4.12}$$

where f' denotes the derivative with respect to t of the function $f(t;\theta_0)$.

Assumption 11 *We assume that the function $f(t;\theta_0)$ is continuously derivable with respect to t, its derivative is denoted as f' and is bounded.*

We will also assume some regularity on the model.

Assumption 12 *$u \to f(u;\theta_0)$ where u ranges over $\theta_0'\mathcal{X}$ is assumed to belong to some Donsker class of functions $\mathcal{F}$.*

In our minds, $\mathcal{F}$ will be the class $\mathcal{C}^1(\theta_0'\mathcal{X}, M)$, that is the class of functions ϕ defined on $\theta_0'\mathcal{X}$ and being one time differentiable with $\|\phi\|_\infty + \|\phi'\|_\infty \leq M$ (see section 2.7 in Van der Vaart and Wellner, 1996). It is important not to impose to much regularity on the regression model, since, as we will see it in Assumption 13, $\hat{f}$ will also be required to belong to this class with probability tending to one.

Assumptions on $\hat{f}$.

Assumption 13 *With probability tending to one, $u \to \hat{f}(u;\theta_0) \in \mathcal{F}$ where $\mathcal{F}$ is defined in Assumption 12. Furthermore,*

$$\|\nabla_\theta \hat{f} - \nabla_\theta f\|_\infty = o_P(1), \tag{4.13}$$

and, defining $W_i^* = \delta_i n^{-1}[1 - G(T_i-)]^{-1}$,

$$\sup_{\theta \in \Theta_n} \left| \sum_{i=1}^n W_i^* J(\theta_0' X_i)[T_i - f(\theta_0' X_i;\theta_0)][\nabla_\theta \hat{f}(X_i;\theta_0) - \nabla_\theta f(X_i;\theta_0)] \right| = o_P(n^{-1/2}), \tag{4.14}$$

$$\sup_{\theta \in \Theta_n} \left| \sum_{i=1}^n W_i^* J(\theta_0' X_i)(\hat{f}(\theta_0' X_i;\theta_0) - f^*(\theta_0' X_i;\theta_0))(\nabla_\theta \hat{f}(X_i;\theta) - \nabla_\theta f^*(X_i;\theta)) \right| = o_P(n^{-1/2}). \tag{4.15}$$

We can now enounce our asymptotic normality theorem.

Theorem 4.2 *Under Assumptions 3 to 13, we have*

$$\begin{aligned}
\sqrt{n}\left(\hat{\theta}^{WLS}-\theta_0\right) &\Rightarrow \mathcal{N}\left(0, V^{-1}W^{WLS}V^{-1}\right), \\
\sqrt{n}\left(\hat{\theta}^{SD}-\theta_0\right) &\Rightarrow \mathcal{N}\left(0, V^{-1}W^{SD}V^{-1}\right).
\end{aligned}$$

Proof. First apply Proposition 6.9 to obtain that $J_n(\theta_n' X_i)$ can be replaced by $J(\theta_0' X_i)$ or by $\mathbf{1}_{f_\theta(\theta' X_i)>c/2}$, plus some arbitrary small terms which will not be mentioned in the following. Moreover, we consider $\theta \in \Theta_n$ which is an $o_P(1)-$neighborhood of θ_0.

Proof for the WLS approach. Using the representation (2.5) of the Kaplan-Meier weights,

$$\begin{aligned}
M_n\left(\theta, \hat{f}\right) &= M_n\left(\theta, f\right) - \frac{2}{n}\sum_{i=1}^{n} \frac{\delta_i J(\theta_0' X_i)\left(T_i - f\left(\theta' X_i;\theta\right)\right)}{1-\hat{G}\left(T_i-\right)} \\
&\quad \times \left[\hat{f}\left(\theta' X_i;\theta\right) - f\left(\theta' X_i;\theta\right)\right] \\
&\quad + \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i J(\theta_0' X_i)}{1-\hat{G}\left(T_i-\right)}\left[\hat{f}\left(\theta' X_i;\theta\right) - f\left(\theta' X_i;\theta\right)\right]^2 \\
&= M_n\left(\theta, f\right) - 2A_{1n} + B_{1n}.
\end{aligned}$$

First decompose A_{1n} into four terms,

$$\begin{aligned}
A_{1n} &= \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i J(\theta_0' X_i)\left(T_i - f\left(\theta_0' X_i;\theta_0\right)\right)}{1-\hat{G}\left(T_i-\right)}\left[\hat{f}\left(\theta_0' X_i;\theta_0\right) - f\left(\theta_0' X_i;\theta_0\right)\right] \\
&\quad + \frac{\delta_i J(\theta_0' X_i)\left(f\left(\theta_0' X_i;\theta_0\right) - f\left(\theta' X_i;\theta\right)\right)}{1-\hat{G}\left(T_i-\right)} \\
&\quad \times \left[\hat{f}\left(\theta' X_i;\theta\right) - f\left(\theta' X_i;\theta\right) - \hat{f}\left(\theta_0' X_i;\theta_0\right) + f\left(\theta_0' X_i;\theta_0\right)\right] \\
&\quad + \frac{\delta_i J(\theta_0' X_i)\left(f\left(\theta_0' X_i;\theta_0\right) - f\left(\theta' X_i;\theta\right)\right)}{1-\hat{G}\left(T_i-\right)}\left[\hat{f}\left(\theta_0' X_i;\theta_0\right) - f\left(\theta_0' X_i;\theta_0\right)\right] \\
&\quad + \frac{\delta_i J(\theta_0' X_i)\left(T_i - f\left(\theta_0' X_i;\theta_0\right)\right)}{1-\hat{G}\left(T_i-\right)} \\
&\quad \times \left[\hat{f}\left(\theta' X_i;\theta\right) - f\left(\theta' X_i;\theta\right) - \hat{f}\left(\theta_0' X_i;\theta_0\right) + f\left(\theta_0' X_i;\theta_0\right)\right] \\
&= A_{2n} + A_{3n} + A_{4n} + A_{5n}.
\end{aligned}$$

A_{2n} does not depend on θ.

For A_{3n}, use Assumption 5 to bound $f(\theta_0' X;\theta_0) - f(\theta' X;\theta)$ by $M \times \|\theta - \theta_0\|$ (for some constant $M > 0$) using a Taylor expansion. Using a Taylor expansion, the bracket in A_{3n}

can be rewritten as $(\theta-\theta_0)'[\nabla_\theta \hat f(X;\tilde\theta)-\nabla_\theta f(X;\tilde\theta)]$ for some $\tilde\theta\in\Theta_n$. Moreover, using Proposition 6.9, we can replace $J(\theta_0'X)$ by $\mathbf{1}_{\{f_{\tilde\theta}(\tilde\theta' X)>c/2\}}$. Hence we have

$$A_{3n} \quad \leq \quad M\|\theta-\theta_0\|^2 \sup_{\theta\in\Theta, x\in\mathcal{X}} |\nabla_\theta \hat f(x;\theta)-\nabla_\theta f(x;\theta)| \int d\hat F_{(X,Y)}(x,y).$$

The uniform consistency of $\nabla_\theta \hat f$ in Assumption 13 shows that $A_{3n}=o_P(\|\theta-\theta_0\|^2)$.

For A_{4n}, use a second order Taylor expansion and the uniform consistency of $\hat f$ to obtain

$$A_{4n} \quad = \quad \frac{1}{n}\sum_{i=1}^n \frac{\delta_i J(\theta_0'X_i)\,(\theta-\theta_0)'}{1-\hat G\,(T_i-)} \nabla_\theta f\,(X_i;\theta_0)\left[\hat f\,(\theta_0'X_i;\theta_0)-f\,(\theta_0'X_i;\theta_0)\right] \quad (4.16)$$

$$+o_P(\|\theta-\theta_0\|^2), \quad (4.17)$$

In the first term, first replace G by $\hat G$. Using Lemma 6.2 ii) with $\eta=1$, this introduces a remainder term which is bounded by

$$\frac{O_P(\|\theta-\theta_0\|n^{-1/2})\|\hat f-f\|_\infty}{n}\sum_{i=1}^n \frac{\delta_i J(\theta_0'X_i)C^{1/2+\varepsilon}(T_i-)}{1-G\,(T_i-)},$$

where we also used the boundedness of $\nabla_\theta f$. Using the uniform consistency of $\hat f$ shows that replacing $\hat G$ by G in (4.16) only arises an $o_P(\|\theta-\theta_0\|n^{-1/2})$ term. Now, we will use the regularity assumption (12) on $f(\cdot;\theta_0)$. If the class $\mathcal{F}$ is Donsker, the class of function $\mathcal{F}'=(\delta,T,X)\to \delta J(\theta_0'X)[1-G(T_i-)]^{-1}\nabla_\theta f(X_i;\theta_0)\mathcal{F}(\theta_0'X_i)$ is Donsker, from a stability property of Donsker classes (see e.g. Van der Vaart and Wellner, 1996). The notation $\mathcal{F}(\theta_0'X_i)$ is used to mention that the functions in $\mathcal{F}$ are evaluated at $\theta_0'X_i$. Furthermore, for all $\phi\in\mathcal{F}'$, $E[\phi(T_i,\delta_i,X_i)]=0$, since

$$E\left[\frac{\delta_i\nabla_\theta f(X_i;\theta_0)}{1-G(T_i-)}|\theta_0'X_i\right]=E\left[\nabla_\theta f(X_i;\theta_0)|\theta_0'X_i\right]=0,$$

from (4.12). Hence, using the fact that $\hat f(\cdot;\theta_0)\in\mathcal{F}$ with probability tending to one, and the asymptotic equicontinuity property of Donsker classes for $\mathcal{F}'$ (see Van der Vaart and Wellner, 1996), we obtain

$$\frac{1}{n}\sum_{i=1}^n \frac{\delta_i J(\theta_0'X_i)\,(\theta-\theta_0)'}{1-G\,(T_i-)} \nabla_\theta f\,(X_i;\theta_0)\left[\hat f\,(\theta_0'X_i;\theta_0)-f\,(\theta_0'X_i;\theta_0)\right]=o_P(\|\theta-\theta_0\|n^{-1/2}),$$

and finally, $A_{4n}=o_P(\|\theta-\theta_0\|n^{-1/2})$.

Similarly, for A_{5n}, a Taylor expansion yields

$$
\begin{aligned}
A_{5n} &= \frac{(\theta-\theta_0)'}{n}\sum_{i=1}^{n}\frac{\delta_i J(\theta_0' X_i)\left(T_i - f\left(\theta_0' X_i;\theta_0\right)\right)}{1-\hat{G}\left(T_i-\right)} \\
&\quad \times\left[\nabla_\theta \hat{f}\left(X_i;\tilde{\theta}\right)-\nabla_\theta f\left(X_i;\tilde{\theta}\right)\right] \\
&= \frac{(\theta-\theta_0)'}{n}\sum_{i=1}^{n}\frac{\delta_i J(\theta_0' X_i)\left(T_i - f\left(\theta_0' X_i;\theta_0\right)\right)}{1-G\left(T_i-\right)} \\
&\quad \times\left[\nabla_\theta \hat{f}\left(X_i;\tilde{\theta}\right)-\nabla_\theta f\left(X_i;\tilde{\theta}\right)\right]+o_P(\|\theta-\theta_0\| n^{-1/2}),
\end{aligned}
$$

where, as for A_{4n}, we replaced $\hat{G}$ by G by using Lemma 6.2 ii) and the uniform consistency of $\nabla_\theta \hat{f}$. Now We then obtain $A_{5n} = o_P(\|\theta-\theta_0\| n^{-1/2}) + o_P(\|\theta-\theta_0\|^2)$ using condition 4.15 in Assumption 13.

For B_{1n}, write

$$
\begin{aligned}
B_{1n} &= \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i J(\theta_0' X_i)}{1-\hat{G}\left(T_i-\right)} \\
&\quad \times\left[\hat{f}\left(\theta' X_i;\theta\right)-f\left(\theta' X_i;\theta\right)-\hat{f}\left(\theta_0' X_i;\theta_0\right)+f\left(\theta_0' X_i;\theta_0\right)\right]^2 \\
&\quad +\frac{\delta_i J(\theta_0' X_i)}{1-\hat{G}\left(T_i-\right)}\left[\hat{f}\left(\theta_0' X_i;\theta_0\right)-f\left(\theta_0' X_i;\theta_0\right)\right] \\
&\quad +\frac{\delta_i J(\theta_0' X_i)}{1-\hat{G}\left(T_i-\right)}\left[\hat{f}\left(\theta_0' X_i;\theta_0\right)-f\left(\theta_0' X_i;\theta_0\right)\right] \\
&\quad \times\left[\hat{f}\left(\theta' X_i;\theta\right)-f\left(\theta' X_i;\theta\right)-\hat{f}\left(\theta_0' X_i;\theta_0\right)+f\left(\theta_0' X_i;\theta_0\right)\right]
\end{aligned}
$$

Using a second order Taylor expansion and arguments similar to those used for A_{3n}, we obtain that the first term is of order $o_P\left(\|\theta-\theta_0\|^2\right)$. The second term does not depend on θ. For the third, a first order Taylor expansion shows that it is bounded by

$$
\|\theta-\theta_0\|\|\nabla_\theta \hat{f}-\nabla_\theta f\|_\infty \sup_{x:J(\theta_0' x)=1}|\hat{f}(\theta_0' x;\theta_0)-f(\theta_0' x;\theta_0)|\int d\hat{F}_{(X,Y)}(x,y).
$$

Now condition 4.14 in Assumption 13 shows that this is $o_P(\|\theta-\theta_0\| n^{-1/2})$.

We have just shown that

$$
M_n\left(\theta,\hat{f}\right)=M_n\left(\theta,f\right)+o_P\left(\frac{\|\theta-\theta_0\|}{\sqrt{n}}\right)+o_P\left(\|\theta-\theta_0\|^2\right),
$$

on a set of probability tending to one. Furthermore, using (4.10) we deduce $\|\theta-\theta_0\| = O_P\left(n^{-1/2}\right)$ from Theorem 1 in Sherman (1994), and since, from (4.11), on $O_P\left(n^{-1/2}\right)$ −neighborhoods of θ_0,

$$
M_n\left(\theta,f_\theta\right)=\frac{1}{2}\left(\theta-\theta_0\right)' V\left(\theta-\theta_0\right)+\frac{1}{\sqrt{n}}\left(\theta-\theta_0\right)' W_n^{WLS}+o_P\left(n^{-1}\right),
$$

we can apply Theorem 2 of Sherman to conclude on the asymptotic law.

Proof for ϕ^{SD}. Proceed as for ϕ^{MC}, the only difference is in the fact that $\hat{G}$ does not appear in the terms where T does not appear at the numerator. ■

5 Simulation study

In this section, we tried to compare the behavior of our estimator with the estimator proposed by Burke and Lu (2005) who used the average derivative technique. We considered three configurations.

Config 1	Config 2	Config 3
$\varepsilon \sim \mathcal{N}(0,2)$	$\varepsilon \sim \mathcal{N}(0,1)$	$\varepsilon \sim \mathcal{N}(0,1/16)$
$X \sim \mathcal{U}[-2;2] \otimes \mathcal{U}[-2;2]$	$X \sim \mathcal{U}[0;1] \otimes \mathcal{U}[0;1]$	$X \sim \mathcal{B}(0.6) \otimes \mathcal{U}[-1;1]$
$f(\theta' x;\theta) = 1/2(\theta' x)^2 + 1$	$f(\theta' x;\theta) = \frac{2e^{(0.5\theta' x)}}{0.5+\theta' x}$	$f(\theta' x;\theta) = 1 + 0.1(\theta' x)^2$
		$-0.2(\theta' x - 1)$
$\theta_0 = (1,1)'$	$\theta_0 = (1,2)'$	$\theta_0 = (1,2)'$
$C \sim \mathcal{U}[0,\lambda_1]$	$C \sim \mathcal{E}(\lambda_2)$	$C \sim \mathcal{E}(\lambda_3)$

The first configuration is used by Burke and Lu (2005) in their simulation study. Observe that, in this model, (2.2) does not hold (this condition (2.2) is also needed in Burke and Lu's approach), but it only introduces some asymptotic bias in the estimation. In the second configuration, there is no such problem since C is exponential. In the third configuration, we see that X does not have a Lebesgue density, but $\theta' X$ does. In this situation, it is expected that the average derivative techniques does not behave well since it requires that X has a density.

In each configuration, we simulated 1000 samples of different size n. For each sample, we computed $\hat{\theta}_{WLS}$, $\hat{\theta}_{SD}$, and $\hat{\theta}_{AD}$ which denotes the average derivative estimate computed from the technique of Burke and Lu (2005). We then evaluated $\|\hat{\theta} - \theta_0\|^2$ for each estimate, in order to estimate the Mean Squared Error (MSE) $E[\|\hat{\theta} - \theta_0\|^2]$. We used different values of the parameters λ_i to modify the proportion of censored responses (15%, 30%, and 50% respectively). Results are presented in the table below.

Config 1		$n = 30$	$n = 50$	$n = 100$
$\lambda_1 = 2.4$	$\hat{\theta}^{AD}$	4.8656×10^{-2}	2.6822×10^{-2}	1.1733×10^{-2}
	$\hat{\theta}^{WLS}$	1.2814×10^{-4}	4.0350×10^{-5}	2.0694×10^{-5}
	$\hat{\theta}^{SD}$	1.2200×10^{-4}	8.3869×10^{-5}	1.3820×10^{-5}
$\lambda_1 = 1.17$	$\hat{\theta}^{AD}$	4.5757×10^{-2}	3.3285×10^{-2}	1.8236×10^{-2}
	$\hat{\theta}^{WLS}$	1.5713×10^{-4}	3.8088×10^{-5}	2.9482×10^{-5}
	$\hat{\theta}^{SD}$	1.6925×10^{-4}	4.0177×10^{-5}	1.9924×10^{-5}
$\lambda_1 = 0.1$	$\hat{\theta}^{AD}$	1.0102×10^{-1}	7.4870×10^{-2}	5.0438×10^{-2}
	$\hat{\theta}^{WLS}$	8.3666×10^{-4}	1.3010×10^{-4}	3.7669×10^{-5}
	$\hat{\theta}^{SD}$	1.2000×10^{-3}	6.7356×10^{-5}	2.3650×10^{-5}
Config 2		$n = 30$	$n = 50$	$n = 100$
$\lambda_2 = 0.2$	$\hat{\theta}^{AD}$	4.1260×10^{-1}	3.6920×10^{-1}	3.4151×10^{-1}
	$\hat{\theta}^{WLS}$	7.8201×10^{-3}	6.5401×10^{-3}	5.8660×10^{-3}
	$\hat{\theta}^{SD}$	1.8296×10^{-2}	1.4721×10^{-2}	1.1034×10^{-2}
$\lambda_2 = 0.1$	$\hat{\theta}^{AD}$	3.5199×10^{-1}	3.3522×10^{-1}	2.8713×10^{-1}
	$\hat{\theta}^{WLS}$	1.2301×10^{-2}	7.8301×10^{-3}	7.7180×10^{-3}
	$\hat{\theta}^{SD}$	2.0822×10^{-2}	2.0301×10^{-2}	1.9741×10^{-2}
$\lambda_2 = 0.05$	$\hat{\theta}^{AD}$	1.6238	1.5553	1.5223
	$\hat{\theta}^{WLS}$	1.6312×10^{-2}	1.5100×10^{-2}	1.2013×10^{-2}
	$\hat{\theta}^{SD}$	3.0344×10^{-2}	2.7057×10^{-2}	2.2510×10^{-2}
Config 3		$n = 30$	$n = 50$	$n = 100$
$\lambda_3 = 11$	$\hat{\theta}^{AD}$	> 10	> 10	> 10
	$\hat{\theta}^{WLS}$	4.1896×10^{-4}	3.1530×10^{-4}	1.7453×10^{-4}
	$\hat{\theta}^{SD}$	4.6218×10^{-4}	1.8696×10^{-4}	1.5286×10^{-4}
$\lambda_3 = 4$	$\hat{\theta}^{AD}$	> 10	> 10	> 10
	$\hat{\theta}^{WLS}$	9.1584×10^{-4}	3.3124×10^{-4}	2.8984×10^{-4}
	$\hat{\theta}^{SD}$	3.4912×10^{-4}	2.3344×10^{-4}	2.2457×10^{-4}
$\lambda_3 = 2$	$\hat{\theta}^{AD}$	> 10	> 10	> 10
	$\hat{\theta}^{WLS}$	2.0159×10^{-2}	1.1431×10^{-2}	2.4111×10^{-4}
	$\hat{\theta}^{SD}$	9.0591×10^{-4}	2.0668×10^{-4}	1.9921×10^{-4}

Globally, the performance of the different estimates shrinks when the proportion of censored responses increases. Performances of $\hat{\theta}^{WLS}$ and $\hat{\theta}^{SD}$ are globally similar. In all tested configurations, $\hat{\theta}^{WLS}$ and $\hat{\theta}^{SD}$ seem to perform better than $\hat{\theta}^{AD}$. As expected, in the situation where X does not have a density, $\hat{\theta}^{AD}$ does not converge.

6 Appendix

6.1 Some results on Kaplan-Meier integrals

In this section, we recall some facts on the behavior of Kaplan-Meier integrals. First part of this section is devoted to the i.i.d representation of Kaplan-Meier integrals derived by Stute (1995, 1996), first in the univariate case, then in presence of covariates. For this, define, for any function ϕ,

$$U_i(\phi) = \int \phi(x,y)\psi(y,T_i,\delta_i)dF_{(X,Y)}(x,y),$$

where ψ has been defined in Theorem 4.1. It can be easily shown that $E[U_i(\phi)] = 0$. The following Theorem has been derived by Stute (1996).

Theorem 6.1 *Let ϕ be a function satisfying*

$$\int |\phi(x,y)| C^{1/2}(y)\, dF_{(X,Y)}(x,y) < \infty.$$

Then

$$\begin{aligned}\int \phi(x,y)\, d\hat{F}_{(X,Y)}(x,y) &= \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i\phi(X_i,T_i)}{1-G(T_i-)} \\ &\quad + \frac{1}{n}\sum_{i=1}^{n} U_i(\phi) + o_P\left(n^{-1/2}\right).\end{aligned}$$

In view of the expression (2.5) of the jumps of Kaplan-Meier estimate, this Theorem shows that, asymptotically, these jumps can be replaced by the "ideal" jumps, say $W_i^* = n^{-1}\delta_i[1-G(T_i-)]^{-1}$, plus some perturbation that only appears in the study of the variance (since its expectation is zero). The following lemma gives some additional precision on the difference between the jumps W_{in} and the "ideal" jumps W_i^*.

Lemma 6.2 *Recall that $\hat{G}$ is the Kaplan-Meier estimator for the distribution of C, $W_{in} = n^{-1}\delta_i[1-\hat{G}(T_i-)]^{-1}$ and $W_i^* = \delta_i[1-G(T_i-)]^{-1}$, and denote by $T_{(n)}$ the largest observation.*

$$\sup_{t\leq T_{(n)}} \frac{1-\hat{G}(t-)}{1-G(t-)} = O_P(1) \quad \text{and} \quad \sup_{t\leq T_{(n)}} \frac{1-G(t-)}{1-\hat{G}(t-)} = O_P(1)\,; \tag{6.18}$$

ii) For all $0 \leq \eta \leq 1$ and $\varepsilon > 0$,

$$|W_{in} - W_i^*| \leq W_i^* \{C\left(T_i\right)\}^{\eta[1/2+\varepsilon]} \times O_P\left(n^{-\eta/2}\right), \tag{6.19}$$

where the $O_P\left(n^{-\eta/2}\right)$ factor does not depend on i.

Proof.

i) The first part of (6.18) follows from Theorem 3.2.4 in Fleming and Harrington (1991). The second part follows for instance as a consequence of Theorem 2.2 in Zhou (1991).

ii) Fix $\eta > 0$ arbitrarily. Since $\int_a^{\tau_H} C^{-1-2\eta}(y)dC(y) < \infty$, for some $a > 0$, apply Theorem 1 in Gill (1983) to see that

$$\sup_{y\leq T_{(n)}} \left[C\left(y\right)\right]^{-1/2-\eta} |Z(y)| = O_P(1), \tag{6.20}$$

where $Z = \sqrt{n}\{\hat{G} - G\}\{1-G\}^{-1}$ is the Kaplan-Meier process. Next, the proof can be completed by using the definitions of W_{in}, W_i^*, property (6.18), and elementary algebra.

■

6.2 Proof of Theorem 4.1

In this section, we show that the criterion M_n^{WLS} and M_n^{SD} satisfy the conditions (4.10) and (4.11). The same properties can be also shown for the synthetic data estimators of Leurgans (1987) and Lai & al. (1995). More precisions can be found in Delecroix & al. (2008). For the sake of simplicity, we only prove it for M_n^{WLS} since the proof for M_n^{SD} uses similar arguments.

Proof for M_n^{WLS}. Write

$$\begin{aligned}
M_n^{WLS}(\theta, f) - M(\theta) &= 2\int (y - f(\theta_0' x; \theta_0)) \{f(\theta_0' x; \theta_0) - f(\theta' x; \theta)\} \\
&\quad \times d(\hat{F}_{(X,Y)} - F_{(X,Y)})(x, y) \\
&\quad + \int \{f(\theta_0' x; \theta_0) - f(\theta' x; \theta)\}^2 \\
&\quad \times d(\hat{F}_{(X,Y)} - F_{(X,Y)})(x, y) \\
&\quad + \int (y - f(\theta_0' x; \theta_0))^2 \, d(\hat{F}_{(X,Y)} - F_{(X,Y)})(x, y). \qquad (6.21)
\end{aligned}$$

The last term does not depend on θ. Let

$$\chi(x, y) = \{y - f(\theta_0' x; \theta_0)\} \nabla_\theta f(x; \theta_0).$$

Using the derivability Assumption 7 and Theorem 6.1, the first term in the right-hand side of (6.21) is

$$\begin{aligned}
& 2(\theta_0 - \theta)' \int \chi(x, y) d\left(\hat{F}_{(X,Y)} - F_{(X,Y)}\right)(x, y) \\
& +2(\theta_0 - \theta)' \left[\int \{y - f(\theta_0' x; \theta_0)\} \nabla_\theta^2 f(x; \tilde{\theta}) \right. \\
& \left. \times\, d(\hat{F}_{(X,Y)} - F_{(X,Y)})(x, y)\right] (\theta_0 - \theta) \\
=\; & 2(\theta_0 - \theta)' \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i \chi(X_i, T_i)}{1 - G(T_i-)} - E\left[\frac{\delta \chi(X, T)}{1 - G(T-)}\right] \right\} \\
& +2(\theta_0 - \theta)' \frac{1}{n} \sum_{i=1}^{n} U_i(\chi) + o_P\left(\|\theta - \theta_0\|^2\right), \qquad (6.22)
\end{aligned}$$

where the o_P-rate comes from the boundedness of $\nabla_\theta^2 f$ and consistency of Kaplan-Meier integrals. Furthermore, the empirical sums in (6.22) weakly converge to centered Gaussian variables at rate $O_P(n^{-1/2})$. For the second term in (6.21), rewrite it as

$$(\theta - \theta_0)' \left[\int \left[\nabla_\theta f\left(\tilde{\theta} x; \tilde{\theta}\right) \nabla_\theta f\left(\tilde{\theta} x; \tilde{\theta}\right)'\right] d(\hat{F}_{(X,Y)} - F_{(X,Y)})(x, y)\right] (\theta - \theta_0).$$

From the boundedness of $\nabla_\theta f$, deduce that this is $o_P\left(\|\theta - \theta_0\|^2\right)$. We thus obtained (4.10). To obtain (4.11), use Theorem 6.1.

6.3 Properties of $\hat{f}$

In this section, we derive some properties of $\hat{f}$ defined by (2.8), and show that this estimate satisfies Assumptions 6 and 13. Our approach consists of comparing $\hat{f}$ to the ideal estimator f^* defined as

$$f^*(\theta' x; \theta) = \frac{\sum_{i=1}^n Y_i^* K\left(\frac{\theta' X_i - \theta' x}{h}\right)}{\sum_{i=1}^n K\left(\frac{\theta' X_i - \theta' x}{h}\right)}, \tag{6.23}$$

that is the estimator based on the true (uncomputable) Y_i^*. Indeed, f^* is a regular kernel estimator based on uncensored variables, and can be studied by traditional nonparametric kernel techniques.

Assumptions on the random variables $X'\theta$.

Assumption 14 *For all* $\theta \in \Theta$, $\theta' X$ *has a density which is continuously derivable, with uniformly bounded derivative.*

Assumptions on the kernel function.

Assumption 15

- *K is symmetric, positive, twice continuously differentiable function with K'' satisfying a Lipschitz condition.*
- $\int K(s)ds = 1$.
- *K has a compact support, say* $[-1; 1]$.

Assumptions on the bandwidth.

Assumption 16

- $nh^8 \to 0$.
- $nh^5[\log(n)]^{-1} = O(1)$.

The first Lemma we propose allows us to obtain uniform convergence rates for the ideal estimator f^* as an immediate corollary.

Lemma 6.3 *Let K be a kernel satisfying Assumption 15. Let $\tilde{K}$ denote either K or its derivative. Let Z be a random variable with 4th order moment, with $m(x) = E[Z|X = x]$ twice continuously differentiable, with derivatives of order 0, 1 and 2 uniformly bounded. Consider, for $d = 0, 1$, and any vectors x and x' in $\mathcal{X}$,*

$$g_n(\theta, x, x', d) = \frac{1}{nh^{1+d}} \sum_{i=1}^n \tilde{K}\left(\frac{\theta' X_i - \theta' x}{h}\right) \left[\tilde{K}\left(\frac{\theta' X_i - \theta' x'}{h}\right)\right]^d Z_i.$$

We have, for $d = 0, 1$,

$$\sup_{\theta,x,x'} |g_n(\theta, x, x', d) - E[g_n(\theta, x, x', d)]| = O_P(n^{-1/2}h^{-[d+1]/2}[\log n]^{1/2}), \quad (6.24)$$

$$\sup_{\theta,x:f_\theta(\theta'x)>c/2} |E[g_n(\theta, x, x, d)] - E[Z|X = x]| = O(h^2), \quad (6.25)$$

$$\sup_{\theta,x,x'} |E[g_n(\theta, x, x, 1)]| = O(1). \quad (6.26)$$

Corollary 6.4 *Under Assumption 15,*

$$\|f^* - f\| = O_P(n^{-1/2}h^{-1/2}[\log n]^{1/2} + h^2),$$

$$\|\nabla_\theta f^* - \nabla_\theta f\| = O_P(n^{-1/2}h^{-3/2}[\log n]^{1/2} + h^2).$$

Proof. For the bias terms (6.25) and (6.26), this can be done by a classical change of variables, a Taylor expansion, and the fact that $\int uK(u)du = 0$ and $\int u^2K(u)du < \infty$.

For (6.24), first consider

$$g_n^{M_n}(\theta, x, x', d) = \frac{1}{nh^{1+d}} \sum_{i=1}^n \tilde{K}\left(\frac{\theta' X_i - \theta' x}{h}\right) \left[\tilde{K}\left(\frac{\theta' X_i - \theta' x'}{h}\right)\right]^d Z_i \mathbf{1}_{Z_i \le M_n}.$$

We then follow the methodology of Einmahl and Mason (2005). From Pakes and Pollard (1989), the family of functions indexed by (θ, x, x', h) (which has a constant envelope function),

$$(X, Z) \to \tilde{K}\left(\frac{\theta' X - \theta' x}{h}\right) \left[\tilde{K}\left(\frac{\theta' X - \theta' x'}{h}\right)\right]^d \mathbf{1}_{Z \le M_n},$$

satisfies the uniform entropy condition of Proposition 1 in Einmahl and Mason (2005) (condition (ii) in their Proposition 1). The other assumptions in their Proposition 1 hold with $\beta = \sigma = \tilde{C}M$, for some constant $\tilde{C}$ not depending on M. We then can apply Talagrand's inequality (see Einmahl and Mason, 2005, and Talagrand, 1994), with $\sigma_{\mathcal{G}}^2 = n^{-1/2}h^{-[d+1]/2}$. Take $M_n = n^{1/2}h^{1/2}$. It follows from Talagrand's inequality that

$$\sup_{\theta,x,x'} |g_n^{M_n}(\theta, x, x', d) - E[g_n^{M_n}(\theta, x, x', d)]| = O_P(n^{-1/2}h^{-[d+1]/2}[\log n]^{1/2}).$$

It remains to consider $g_n^{M_n} - g_n$. This difference is bounded by $\tilde{C}n^{-1}h^{-[1+d]} \sum_{i=1}^n |Z_i| \mathbf{1}_{Z_i \ge M_n}$ for some constant $\tilde{C}$. This is a sum of positive quantities, thus we only have to show that its expectation is $o_P(n^{-1/2}h^{-[d+1]/2}[\log n]^{1/2})$. For this, apply Hölder inequality to bound this expectation by $h^{-[1+d]}E[Z^4]^{1/4}\mathbb{P}(Z \ge M_n)^{3/4}$. Moreover, $\mathbb{P}(Z \ge M_n) \le E[Z^4]/M_n^4$ from Tschebychev inequality, and the result follows. ∎

Proposition 6.5 below ensures that the difference between $\hat{f}$ and f^*, in view of uniform consistency, is asymptotically negligible. Hence Assumption 6 can be deduced from the uniform consistency of f^*.

Proposition 6.5 *Under Assumptions 10, 14, and Kernel Assumptions 15 and 16, we have* $\|\hat{f} - f^*\|_\infty + \|\nabla_\theta \hat{f} - \nabla_\theta f^*\|_\infty = o_P(1)$.

Corollary 6.6 *Under the Assumptions of Proposition 6.5, $\hat{f}$ satisfies Assumption 6 and condition (4.13) in Assumption 13.*

Proof. Let $\hat{f}_{\theta' X}(u) = n^{-1}h^{-1}\sum_{i=1}^n K((\theta' X_i - u)/h)$. We have

$$\hat{f}(u;\theta) - f^*(u;\theta) \quad = \quad \frac{1}{h}\sum_{i=1}^n \frac{[W_{in} - W_i^*]T_i K\left(\frac{\theta' X_i - \theta' x}{h}\right)}{\hat{f}_{\theta' X}(u)}. \tag{6.27}$$

Now, from uniform consistency of kernel density estimator (see, e.g. Einmahl and Mason, 2005),

$$\sup_{x\in\mathcal{X},\theta\in\Theta} |\hat{f}_{\theta' X}(\theta' x) - f_{\theta' X}(\theta' x)| = o_P(1).$$

Using this result on the set $\{f_{\theta' X}(\theta' x) > c > 0\}$, and Lemma 6.2 ii) with η sufficiently small, we obtain the bound

$$|\hat{f}(\theta' x;\theta) - f(\theta' x;\theta)| \leq O_P(n^{-\eta/2}h^{-1}) \times \sum_{i=1}^n W_i^* T_i C^{\eta(1/2+\varepsilon)}(T_i -) K\left(\frac{\theta' X_i - \theta' x}{h}\right), \tag{6.28}$$

where the O_P-rate does not depend on θ nor x. Recalling the definition of W_i^*, consider the family of functions indexed by θ and x,

$$\{(T,\delta,X) \rightarrow \delta T[1 - G(T-)]^{-1} C^{\eta(1/2+\varepsilon)}(T-) K((\theta' X - \theta' x)/h)\}.$$

This family is Euclidean (see Lemma 22 in Nolan and Pollard, 1987) for an enveloppe $\delta T[1 - G(T-)]^{-1} C^{\eta(1/2+\varepsilon)}(T-)$ which is, for $\eta = 1/2$, square integrable from Assumption 10. Therefore, using the assumptions on the bandwidth,

$$\sup_{x\in\mathcal{X},\theta\in\Theta} |\sum_{i=1}^n W_i^* T_i C^{\eta(1/2+\varepsilon)}(T_i -) K\left(\frac{\theta' X_i - \theta' x}{h}\right)| = O_P(h) + O_P(n^{-1/2}).$$

Finally, back to (6.28), this shows that $\|\hat{f} - f\|_\infty = O_P(n^{-1/4}) = o_P(1)$.

Similarly, $\|\nabla_\theta \hat{f} - \nabla_\theta f\|_\infty = O_P(n^{-1/4}h^{-1})$.

Now, to prove the corollary, we have to show the uniform consistency of f^* and $\nabla_\theta f^*$, which can be done applying Theorem 2 in Einmahl and Mason (2005). ■

The following Proposition allows us to obtain that $\hat{f}$ satisfies conditions (4.14) and (4.15) of Assumption 13.

Proposition 6.7 *Let $\|\cdot\|_{\Theta_n}$ denote the supremum of the absolute value over Θ_n. Under the Assumptions of Proposition 6.5, we have*

$$h^2 \left\| \sum_{i=1}^n W_i^* J(\theta_0' X_i)(T_i - f(\theta_0' X_i;\theta_0))(\nabla_\theta \hat{f}(X_i;\theta) - \nabla_\theta f^*(X_i;\theta)) \right\|_{\Theta_n} = O_P(n^{-1}), \tag{6.29}$$

$$h^2 \left\| \sum_{i=1}^n W_i^* J(\theta_0' X_i)(\hat{f}(\theta_0' X_i;\theta_0) - f^*(\theta_0' X_i;\theta_0))(\nabla_\theta \hat{f}(X_i;\theta) - \nabla_\theta f^*(X_i;\theta)) \right\|_{\Theta_n} = O_P(n^{-1}), \tag{6.30}$$

$$\left\| \sum_{i=1}^n W_i^* J(\theta_0' X_i)(\hat{f}(\theta_0' X_i;\theta_0) - f^*(\theta_0' X_i;\theta_0))(\nabla_\theta f(X_i;\theta) - \nabla_\theta f^*(X_i;\theta)) \right\|_{\Theta_n} = o_P(n^{-1/2}), \tag{6.31}$$

$$\left\| \sum_{i=1}^n W_i^* (f(\theta_0' X_i;\theta_0) - f^*(\theta_0' X_i;\theta_0))(\nabla_\theta \hat{f}(X_i;\theta) - \nabla_\theta f^*(X_i;\theta)) \right\|_{\Theta_n} = o_P(n^{-1/2}). \tag{6.32}$$

Corollary 6.8 *Under the assumptions of Proposition 6.5, $\hat{f}$ satisfies conditions (4.14) and (4.15) of Assumption 13.*

Proof of Corollary 6.8. To prove (4.15), according to Proposition 6.7, it remains to show that

$$\sup_{\theta\in\Theta_n} \left| \sum_{i=1}^n W_i^* J(\theta_0' X_i)(f(\theta_0' X_i;\theta_0) - f^*(\theta_0' X_i;\theta_0))(\nabla_\theta f(X_i;\theta) - \nabla_\theta f^*(X_i;\theta)) \right| = o_P(n^{-1/2}),$$

which can be done following the lines of Lemma C2 in Delecroix & al. (2008). Similarly, Proposition (6.7) allows to replace $\hat{f}$ by f^*. ■

Proof of Proposition 6.7. We only prove (6.30) and (6.32) since the others are similar.

We first prove (6.30). This can be done by studying separately the different terms arising by differentiation with respect to θ in the definition of $\hat{f}$. We will only study the term coming from the differentiation of the numerator (since the other is similar), that is

$$\frac{1}{nh^2}\sum_{i,j}\frac{\delta_i J(\theta_0'X_i)(\hat{f}(\theta_0'X_i;\theta_0)-f^*(\theta_0'X_i;\theta_0))}{1-G(T_i-)}K'\left(\frac{\theta'X_i-\theta'X_j}{h}\right)\hat{f}_{\theta'X}^{-1}(\theta'X_i)[W_j^*-W_{jn}]T_j.$$

By bounding $|K'|$ by $\|K'\|_\infty$ and using the convergence rate of f^*, it is easily seen that the terms for $i=j$ can be removed from this double sum, arising an $o_P(n^{-1/2})$ term uniform in θ. Applying (6.27), we then get that the above quantity is, up to an $o_P(n^{-1/2})$term,

$$\frac{1}{h^3}\sum_{i\neq j,k}W_i^*J(\theta_0'X_i)K\left(\frac{\theta_0'X_i-\theta_0'X_k}{h}\right)[W_k^*-W_{kn}]T_k$$
$$\times K'\left(\frac{\theta'X_i-\theta'X_j}{h}\right)[W_j^*-W_{jn}]T_j\hat{f}_{\theta'X}(\theta'X_i)^{-1}\hat{f}_{\theta_0'X}(\theta_0'X_i)^{-1}.$$

Again, using Lemma 6.2 ii) with $\eta=1/2$, and bounding K' by $\|K'\|_\infty$ allows us to remove the terms for $j=k$ and $i=k$. For the rest of this triple sum, apply Lemma 6.2 ii) with $\eta=1$ and bound K' by $\|K'\|_\infty$. If follows that the left-hand side of (6.30) is bounded, uniformly in θ, by

$$\frac{O_P(n^{-1}h^{-2})}{n}\sum_{i\neq k,j\neq k,i\neq j}W_j^*W_k^*W_i^*C^{1/2+\varepsilon}(T_j-)C^{1/2+\varepsilon}(T_k-)|T_j||T_k|K\left(\frac{\theta_0'X_i-\theta_0'X_k}{h}\right).$$

The last sum as finite expectation (and does not depend on θ) from Assumption 10.

For (6.32), again, we will consider only the part of $\nabla_\theta\hat{f}$ coming from the differentiation of the numerator, this means that we are trying to bound

$$\frac{1}{h^2}\sum_{i,j}W_i^*J(\theta_0'X_i)[f(\theta_0'X_i;\theta_0)-f^*(\theta_0'X_i;\theta_0)]K'\left(\frac{\theta'X_i-\theta'X_j}{h}\right)T_j[W_j^*-W_{jn}]\hat{f}_{\theta'X}(\theta'X_i)^{-1}. \tag{6.33}$$

First, let S_τ be the double sum deduced from (6.33) by introducing $\mathbf{1}_{T_j\leq\tau}$ for some $\tau<\tau_H$. From Gill (1983), $\sup_{t\leq\tau}|\hat{G}(t)-G(t)||1-G(t)|^{-1}=O_P(n^{-1/2})$, and consequently, $\sup_j|W_j^*-W_{jn}|\mathbf{1}_{T_j\leq\tau}=O_P(n^{-1/2})$. Now, using the uniform convergence rate of f^* and bounding K' by $\|K'\|_\infty$ shows that $S_\tau=o_P(n^{-1/2})$ for any $\tau<\tau_H$. To obtain a bound for (6.33), we then have to make τ tend to τ_H. For this, we use the same Cramer-Slutsky argument as Stute (1995) in his proof of the Central Limit Theorem under censoring.

Using Lemma 6.2 ii) with $\eta=1$, observe that

$$|S_{\tau_H}-S_\tau|\leq O_P(n^{-1/2}h^{-2})\|f^*-f\|_\infty\frac{1}{h}\sum_{i=1}^n K\left(\frac{\theta_0'X_i-\theta_0'X_j}{h}\right)\mathbf{1}_{T_j\geq\tau}W_j^*C^{1/2+\varepsilon}(T_j-)W_i^*.$$

The last part does not depend on θ and its expectation tends to zero as $\tau \to \tau_H$, while the rest is $O_P(n^{-1/2})$, using the convergence rate of f^* and the Assumptions 16. Then the Cramer-Slutsky argument of Stute allows us to conclude. ■

The only condition that still needs to be checked is that $u \to \hat{f}(u;\theta_0) \in \mathcal{F}$, where $\mathcal{F}$ is defined in Assumption 12. This can be done if we specify this class of functions. If $\mathcal{F} = \mathcal{C}^1(\theta_0' X, M)$, it suffices to show that $\sup_u |\hat{f}'(u;\theta_0) - f'(u;\theta_0)| = o_P(1)$, which can be done by using the same method as in Proposition 6.7 to replace f by f^*.

6.4 Trimming

In the following proposition, we show that the trimming $J_n(\theta_n' x)$ can be replaced by $J(\theta_0' x)$ modulo arbitrary small terms.

Proposition 6.9 *Let, for any function ϕ,*

$$R_n = \frac{1}{n}\sum_{i=1}^{n} \phi(\theta, \hat{G}, \hat{f}; T_i, \delta_i, X_i) \left[J(\theta_0' X_i) - J_n(\theta_n' X_i) \right].$$

We have $R_n = o_P\left(\frac{1}{n}\sum_{i=1}^{n} \phi(\theta, \hat{G}, \hat{f}; T_i, \delta_i, X_i)\right) o_P(n^{-1/2})$.

Proof. For any $\delta > 0$, we have, with probability tending to one,

$$|J(\theta_0' X_i) - J_n(\theta_n' X_i)| \leq \mathbf{1}_{f_{\theta_0' X}(\theta_0' x) \leq c - \delta, \hat{f}_{\theta_n' X}(\theta_n' X) \geq c} + \mathbf{1}_{[\delta;\infty]}(Z_n),$$

where $Z_n = \sup_x |\hat{f}_{\theta_n' X}(\theta_n' X) - f_{\theta_0' X}(\theta_0' x)| \tilde{J}(x)$. As in Delecroix, Hristache, Patilea (2006) page 737-738, we have

$$R_n = o_P\left(\frac{1}{n}\sum_{i=1}^{n} \phi(\theta, \hat{G}, \hat{f}; T_i, \delta_i, X_i)\right) + \mathbf{1}_{[\delta;\infty]}(Z_n) \times O_P(1).$$

Note that $\mathbb{P}(n^{1/2} Z_n \geq \delta) \leq \mathbb{P}(Z_n \geq \delta)$, which tends to zero as δ tends to zero. ■

References

[1] Andersen, P. K. & Gill, R. D. (1982) Cox's Regression Model for Counting Processes : A Large Sample Study. *Ann. Statist.* **82**, 1100–1120.

[2] Buckley, J., and James, I. R. (1979). Linear regression with censored data. *Biometrika* **66**, 429–436.

[3] Burke, M. D., and Lu, X. (2005). Censored multiple regression by the method of average derivatives. *J. Multivariate Anal.* **95**, 182–205.

[4] Csörgő, S. (1996). Universal Gaussian approximations under random censorship. *Ann. Statist.* **24**, 2744–2778.

[5] Delecroix, M., Hristache, M., and Patilea, V. (2006). On semiparametric M-estimation in single-index regression. *Journal of Statistical Planning and Inference* **136**, 730–769.

[6] Delecroix, M., Lopez, O., and Patilea, V. (2008)...

[7] Dominitz, J., and Sherman, R. P. (2003). Some convergence theory for iterative estimation procedures. *Econometric Theory* **21**, 838–864.

[8] Einmahl, U. & Mason, D. M., (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.*, **33**, 1380–1403.

[9] Fan, J., and Gijbels, I. (1994). Censored regression: local linear approximations and their applications. *J. Amer. Statist. Assoc.* **89**, 560–570.

[10] Gill, R. (1980). *Censoring and Stochastic Integrals.* Mathematical Centre Tracts **124**. Mathematisch Centrum, Amsterdam.

[11] Gill, R. (1983). Large Sample Behaviour of the Product-Limit Estimator on the Whole Line. *Ann. Statist.* **11**, 49–58.

[12] Härdle, W., and Stoker, T. M. (1989). Investigating Smooth Multiple Regression by the Method of Average Derivatives. *J. Amer. Stat. Ass.* **84**, 986–995.

[13] Heuchenne, C., and Van Keilegom, I. (2005). Estimation in nonparametric location-scale regression models with ensored data. Discussion Paper (DP 0518), Institute of Statistics, Louvain-la-Neuve.

[14] Heuchenne, C., and Van Keilegom, I. (2006). Polynomial regression with censored data based on preliminary nonparametric estimation. *Ann. Inst. Statist. Math.* (to appear).

[15] Ichimura, H. (1993). Semiparametric least squares (SLS)and weighted SLS estimation of single-index models. *Journal of Econometrics* **58**, 71–120.

[16] Koul, H., Susarla, V., and Van Ryzin, J. (1981). Regression analysis with randomly right censored data, *Ann. Statist.* **9**, 1276–1288.

[17] Lai, T. L., Ying, Z., (1991) Large sample theory of a modified Buckley-James estimator for regression analysis with censored data. *Ann. Statist.* **19**, 1370–1402.

[18] Lai, T. L., Ying, Z., and Zheng, Z. (1995) Asymptotic Normality of a Class of Adaptive Statistics with Applications to Synthetic Data Methods for Censored Regression. *J. Multivariate Anal.* **52**, 259–279.

[19] Leurgans, S. (1987). Linear models, random censoring and synthetic data. *Biometrika* **74**, 301–309.

[20] Nolan, D. and Pollard, D. (1987). U-processes : rates of convergence. *Ann. Statist.* **15**, 780–799.

[21] Pakes, A., and Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica* **57**, 1027–1057.

[22] Ritov, Y. (1990). Estimation in a Linear Regression Model with Censored Data. *Ann. Statist.* **18**, 303–328.

[23] Satten, G. A., and Datta, S. (2001). The Kaplan-Meier estimator as an inverse-probability-of-censoring weighted average. *Amer. Statist.* **55**, 207–210.

[24] Sheehy, A., and Wellner, J. A. (1992). Uniform Donsker Classes of functions. *Ann. Probab.* **20**, 1983–2030.

[25] Sherman, R. P. (1994). Maximal inequalities for degenerate U-processes with applications to optimization estimators. *Ann. Statist.* **22**, 439–459.

[26] Shorack, G. R., and Wellner, J. A. (1986). *Empirical processes with applications to statistics.* Wiley, New York.

[27] Stute, W. (1993). Consistent estimation under random censorship when covariables are present. *J. Multivariate Anal.* **45**, 89–103.

[28] Stute, W., and Wang, J.-L. (1993) The strong law under random censorship. *Ann. Statist.* **21**, 1591–1607.

[29] Stute, W. (1995). The central limit theorem under random censorship. *Ann. Statist.* **23**, 422–439.

[30] Stute, W. (1996). Distributional convergence under random censorship when covariables are present. *Scand. J. Statist.* **23**, 461–471.

[31] Stute, W. (1999). Nonlinear censored regression. *Statistica Sinica* **9**, 1089–1102.

[32] Van der Vaart, A.W. (1996). Other Donsker classes. *Ann. Probab.* **24**, 2128–2140.

[33] Van der Vaart, A.W., and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes.* Springer-Verlag, New-York.

[34] Zhou, M. (1992a). M-estimation in censored linear models. *Biometrika* **79**, 837–841.

[35] Zhou, M. (1992b). Asymptotic normality of the "synthetic data" regression estimator for censored survival data. *Ann. Statist.* **20**, 1002–1021.